Service Differentiation in Cognitive Radio Networks: A Priority Queueing Model

Arash Azarfar, Jean-François Frigon and Brunilde Sansò
Department of Electrical Engineering
École Polytechnique de Montréal, Montréal, QC, H3C 3A7, Canada
Email: {arash.azarfar, j-f.frigon, brunilde.sanso} @polymtl.ca

***Abstract*—The popularity of opportunistic spectrum access (OSA), implemented by cognitive radios (CRs), necessitates that these networks should be able to provide service differentiation for different classes of traffic. One of the classic schemes for such a goal is through priority queueing. As link interruptions in OSA networks are frequent, we discuss in this paper four priority queueing disciplines in the presence of interruptions: preemptive-resume, non-preemptive, exceptional non-preemptive and preemptive in case of occurrence of an interruption. Analytical results, in addition to simulation results to validate their accuracy, are provided.**

I. Introduction

The deployment of wireless services in diverse applications and networks implies that wireless links may carry different types of traffic with variable requirements and importance. Meanwhile, it is expected that opportunistic spectrum access (OSA) implemented by Cognitive Radio (CR), due to its comprehensive and powerful features, will be widely deployed in next generation wireless networks [1], [2]. Thus, as illustrated in Figure 1, a cognitive radio link may carry different classes of traffic where it is obliged to frequently release the channel to primary users.

It is therefore crucial to develop a model which enables us to analyze the traffic metrics, such as throughput and delay, for differentiated services in CR networks where the operating channel is experiencing frequent failures due to the appearance of primary users or quality variations.

One of the classic schemes to provide service differentiation in communication networks is through priority queueing, which is well discussed in the literature. However, in CR networks, the channel (i.e., the server of the queue) is subject to interruptions thus priority queueing should be studied in the context of a queueing model with server interruptions. In this paper, we introduce a priority queueing approach for a CR link with multiple classes of traffic, where the link is subject to recurrent failures and interruptions. The occurrence of interruptions represent the time where the operating channel becomes occupied by primary users or its quality is unacceptable due to other reasons such as deep fading. The motivation for this work is that not only classic disciplines of preemptive-resume and non-preemptive, which are defined based on the interaction of different classes of traffic, should be revisited in the presence of interruptions, new priority disciplines can also be defined considering the occurrence of interruptions. These new disciplines take into account also the interaction of interruptions and different traffic classes. We discuss two such queueing disciplines in this paper; so totally, four priority disciplines will be investigated:

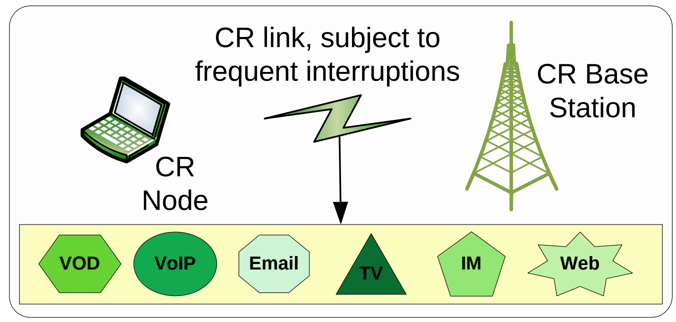

Fig. 1. A CR link with multiple classes of traffic.

- *Non-preemptive* (Non) where the server can not be preempted when any traffic is in service (even in the first arrival recovery period);
- *Exceptional Non-preemptive* where the server can not be preempted in general, but it can be preempted in the arrival recovery period, before the start of the real service of low priority packets;
- *Preemptive* (Pr) where a high-priority class can preempt the server any time;
- *Preemptive in case of failure* where the server can not be preempted during the service, but if an interruption occurs, after the interruption, the high priority packet (if any) will be served first. This scheme, to the best of our knowledge, is a new scheme which is very well applicable to CR networks.

To tackle these priority queueing disciplines, we first discussed in [3] a general queueing model (with single class of CR traffic) for a CR link where the link was subject to interruptions. The results of that work will be extended here to priority queueing for multiple classes of CR traffic. Note that even the results for the CR queue with a single class of traffic, which was discussed in [3] for limited cases and are recalled here with more discussions, are more general compared to other papers who model the single-traffic CR queue with interruptions as a priority queue with primary and secondary users [4], [5]. The reason is twofold: First, priority queueing models will be very complicated when the distribution of the duration of the primary users' busy periods

are general. Second and more important is that, interruptions in our model may represent the intervals in which, the user is performing spectrum handover (multi-channel scenarios), not necessarily the busy periods of the primary users. Although previous models can provide accurate results for simple cases, they do not address the general interaction between primary and secondary users which can limit their usefulness. For instance, their results can not generally be used to discuss the relation between different CR traffic classes, which is the objective of this paper.

In the literature, the notions of queueing with interruption and priority queueing have been very well discussed [6]–[10]. In these models, the interruption periods are the busy periods generated by higher priority of traffic and the priority discipline is preemptive. However in this paper, we consider a general distribution for the interruptions and also we discuss other priority disciplines such as non-preemptive (Non), exceptional non-preemptive (ENo) and preemptive in case of failure (FP).

The results of this paper provide utility relations for the purpose of differentiated packet-level performance evaluation in cognitive radio networks. The interruption periods can model the time intervals in which the CR network is searching for a new channel. In this case, these periods can be called recovery periods [11]. We focus on this model thus recovery and interruption periods are used interchangeably. Any new proposed recovery scheme in physical or MAC layer can be easily compared with previous ones just by deriving the recovery time distribution and changing the parameters of the interruptions in the queueing model.

Moreover, derived relations can be employed in optimization models to adjust the performance of the traffic. In CR networks which are equipped to learning and decision-making capabilities, this aspect could be very interesting. For instance, proposed priority schemes can be employed in an optimization and decision-making problem where the objective is optimizing some performance metrics for different classes of traffic, where the decision to be made periodically is the priority queueing scheme to be selected among these four proposed disciplines.

The main contributions of this paper can be listed as follows:

- A comprehensive queueing model for opportunistic spectrum access (OSA);
- Analytical utility relations for four priority queueing disciplines with interruptions;
- A new priority queueing discipline which is preemption in case of failure;

The reminder of the paper is organized as follows. In Section II, an M/G/1/ queue with interruptions and with a single class of CR traffic is discussed. The results are then used in Section III to solve four priority queueing disciplines in the presence of interruptions. Simulation results are presented in Section IV and finally Section V concludes the paper with some remarks on future research directions.

II. General Queue (Single Traffic Class) Analysis

Consider a Cognitive Radio (CR) node (user) operating in a CR network using opportunistic spectrum access (OSA). This user continues operating over a channel, called its current operating channel, until it finds the channel unavailable due to the return of primary users or with an unacceptable transmission quality due to deep fading, for example. In a multi-channel scenario, the CR user starts a recovery process with a random duration to find a new channel and then restarts the operation over this new channel. For a single-channel scenario, the CR user waits for the channel to become available again. By modeling the event of missing or vacating the channel as a failure [11], we can denote the operating (availability) and recovery periods as *time to failure* and *time to recovery*, respectively.

Let two random variables Y and R represent the length of the operating and recovery periods with cumulative distribution functions F_Y and F_R, respectively. As illustrated in Figure 2, the operation of this user can be modeled with alternating renewal processes, assuming homogeneous channels. From the queueing point of view, the output buffer of the user can be modeled as a queue with random service interruptions. Random interruption implies that both the time of the failure occurrence which triggers a recovery, and the duration of the interruption (recovery time or waiting time) are random variables. If we assume that the arrival process of all classes of traffic follows a Poisson distribution with a general service time distribution, the operation of the node can be modeled as an M/G/1 queue with random service interruptions.

As can be seen in Figure 2, we are now ready to mathematically tackle the queueing model where the random variable Y represents the availability (ready for use) periods and R stands for the interruption (recovery) periods. Let us define also a new random variable C which is the sum of two random variables Y and R and represents the length of a cycle made of alternating renewal processes Y and R. The distribution of C can be given from the convolution of the distributions of Y and R.

The service rate is fixed, so the distribution of the packet length defines the distribution of the real service time of the packets. The fixed service rate may imply however a fast varying channel whose average service rate is used as a fixed service rate in this model. The random variable T is used to represent the real service time (transmission time) of the packets. We call it *real service time* because the time that a packet spends in service in such a queue with interruptions includes both the real service time and the interruptions that may occur during its service. The notion of *completion time* is used for the whole time in service. λ represents the Poisson arrival rate. The queue size is assumed infinite, so the main performance metric is the total time spent in the system (called system time or sojourn time) and in the queue (queueing time), represented by D and W respectively. Throughout the paper, for any random variable Z, $f_Z(.)$ and $F_Z(.)$ respectively represent the probability density/mass function (PDF or PMF) and the cumulative distribution function (CDF) of the random variable Z. Moreover, $\hat{Z}(s)$ represents the Laplace-Stieltjes (LST) of the distribution of the random variable Z ($F_Z(.)$). All the parameters and notations are listed in Table II.

As can be seen in Figures 3 and 4 and discussed in [3], we can distinguish three types of packet in such a queueing

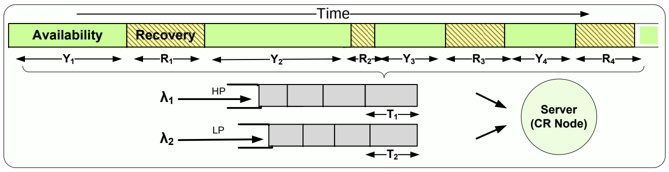

Fig. 2. Queue Model with Interruption.

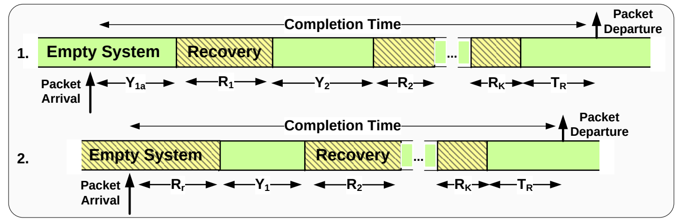

Fig. 3. Completion time for the case ‘a’ (1.) and case ‘u’ (2.).

model:

1) When the packet enters an empty system and the server (i.e., current channel) is available, its real service starts immediately (See Figure 3.1. The index ‘a’ is used to designate this case);
2) When the packet enters an empty system during an interruption period such that the server is not available and a recovery before starting serving the packet needs to be completed (See Figure 3.2. The index ‘u’ is used to designate this case);
3) When the packet enters a busy system and is queued, its service starts immediately after the completion time of the previous packet (See Figure 4. The index ‘b’ is used to designate this case). Note that in this case, the completion time of the packet is always started within an availability period (similar to the case ‘a’).

As there are two perspectives for queueing with interruptions, we discuss both in the beginning. Their difference is for the packets which arrive during an empty system and the system is in the recovery period (case ’u’).

In the first model, the arrival time is the start of the completion time in the case ‘u’ (Figure 3.2). However, the real service time of a packet is started when the current recovery period is completed. So, the remaining time of the current recovery period (R_r) is not accounted as waiting time, but the completion time.

In the second model which is used more in the literature, the start of the completion time is always the start of the real service. Therefore, the remaining time of the recovery period in arrival for the case ’u’ (R_r) is considered as waiting time. To distinguish these two, we use a star, ’*’, for the first case. Note that as expected, the system time for both cases will be the same and the difference is for the completion time and waiting time of the case ‘u’ packets.

It is worth mentioning that when the transmission of a packet is interrupted, the whole packet should be generally re-transmitted and the assumption of a partial transmission (resume of transmission) is not realistic in some applications. This case will be postponed for future and we discuss here only the cases that partial transmission is possible and thus after an interruption, the node transmits the remaining part of the packet (service-resume).

A. Case ’a’: Arrival to an Empty-available System

For packets of case ’a’, the waiting time is zero. As the first availability period is started in a random point during an availability period, we assumed that the remaining time until the next interruption is distributed with a random variable called Y_{1a} (first availability period for case ’a’). As illustrated in Figure 3.1, the completion time of the packet in the first case, X_a, can be given by:

$$X_a = \begin{cases} T, & Y_{1a} \geq T \\ Y_{1a} + R_1 + Y_2 + R_2 + ... + Y_K + R_K + T_R, & \text{Otherwise.} \end{cases} \quad (1)$$

where T_R stands for the real transmission time of the last part of the packet in the last operating period. For a service time distributed with the random variable T, we have:

$$T = Y_{1a} + ... + Y_K + T_R. \quad (2)$$

That is, the sum of transmission times in each availability period should be equal to T. The unknown parameter is K which is the number of availability periods during the completion time. If we consider only the operating periods, Y_i, as a renewal process, K in all scenarios is the number of renewals of Y during the real transmission time of a packet where its distribution can be found from the renewal theory results [12], [13].

The exact distribution of Y_{1a} depends on the time that the user has switched to this channel, i.e., after the last recovery period (multi-channel scenarios), or the time that the system has become empty. Finding the length of this time interval can be complicated for general distributions. If we find the distribution of Y_{1a}, the moments of the number of renewals, K, can be given by:

$$\mathcal{L}[m_a(t)] = \frac{\hat{Y_{1a}}(s)}{s(1-\hat{Y}(s))} \quad (3)$$

where $m_a(t)$ is the mean renewal function (average number of renewals during $(0, t]$) composed of instances of Y_{1a} and Y [12], [13]. One can find $m_a(t)$ from the equation above and thereby derive $E[K_a]|(T = t)$ [1]. Unconditional mean number of renewals can then be given by:

$$E[K_a] = \int_0^\infty m_a(t) f_T(t) dt \quad (4)$$

Now let $m_a^2(t)$ be the second moment of the number of renewals in time t. The LST of $m_u^2(t)$ can be found as:

$$\mathcal{L}[m_a^2(t)] = \frac{\hat{Y_{1a}}(s)(1+\hat{Y}(s))}{s\left(1-\hat{Y}(s)\right)^2}, \quad (5)$$

so that $E[K_a^2]|(T = t)$ is found.

[1] K_a : K for the case ‘a’

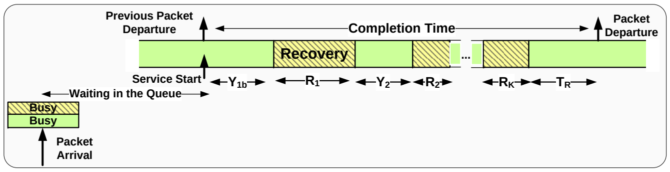

Fig. 4. Completion time for the third case when the packet enters a busy system and is queued (case 'b').

When the first two moments of K_a is found, using Eq. (1), we can give the first moment of X_a by:

$$E[X_a] = E[T] + E[K_a]E[R]. \tag{6}$$

The second moment can similarly be found equal to:

$$E[X_a^2] = E[T^2] + 2E[TK_a]E[R] + E[K_a](E[R^2] - (E[R])^2) + (E[R])^2 E[K_a^2], \tag{7}$$

using the fact that the number of renewals, K_a, is independent of the recovery process, but not the service time. The second term in the calculation of $E[X_a]$ is a random sum whose moments can be found easily [2]. $E[TK_a]$ can be found by conditioning on the value of T, as follows:

$$\begin{aligned} E[TK_a] &= \int_0^\infty E[TK_a]|(T=t) f_T(t)dt \\ &= \int_0^\infty t m_a(t) f_T(t)dt. \end{aligned} \tag{8}$$

B. Case 'b': Arrival to a busy system

If we find the number of renewals for the case 'b' equal to K_b, based on the distribution of the first availability period which is here distributed with another distribution, Y_{1b} (Figure 4), we can use all previous results just by replacing Y_{1a} with Y_{1b} and K_a with K_b. However in this case, the waiting time is not zero.

C. Case 'u': Arrival to an empty-unavailable system

For the case 'u', as we discussed, we have two models. In the first model, the completion time of the user is started within a recovery period. As illustrated in Figure 3.2, we have:

$$\begin{aligned} X_u^* &= R_r + Y_1 + R_2 + ... + Y_K + R_{K+1} + T_R \\ X_u &= Y_1 + R_2 + ... + Y_K + R_{K+1} + T_R, \end{aligned} \tag{9}$$

and

$$\begin{aligned} W_u^* &= 0, \\ W_u &= R_r. \end{aligned} \tag{10}$$

The condition is similar to Eq. (2), replacing Y_{1a} with Y_1. Similar to the discussion for the first operating period in the previous case, here the distribution of the first recovery time (R_r) (remaining time of R) can be different from other interruption intervals because the server knows how long it has spent so far for the recovery.

[2]For instance, $Var(KR) = E[K]Var(R) + (E[R])^2 Var(K)$

The number of renewals is independent of the model and it can be found by replacing Y_{1a} in Eqs. (3) to (5) by Y because here the first availability period is also distributed as Y. The moments of the completion time can be found similarly with a small difference that we should add the first recovery period (with distribution R_r) to the completion time of the first model. In other words, we have:

$$\begin{aligned} E[X_u^*] &= E[R_r] + E[X_u], \\ E[X_u^{2*}] &= E[R_r^2] + E[X_u^2] + 2E[R_r]E[X_u]. \end{aligned} \tag{11}$$

As discussed, we can write:

$$\begin{aligned} X_u &= X_a \text{ when } {}_{(K_a \to K, Y_{1a} \to Y)} \\ &= X_b \text{ when } {}_{(K_b \to K, Y_{1b} \to Y)}. \end{aligned} \tag{12}$$

The distribution of R_r can be written as:

$$R_r = R - A|(R > A), \tag{13}$$

where A is the inter-arrival time. That is, the time to arrival of the first packet is an inter-arrival time but conditioned on the fact that it should be less than R. The moments of R_r based on R and A are found in Appendix.

D. Overall Performance Metrics

The first moment of the general completion time can be given by solving the following equation:

$$E[X^*] = \rho^* E[X_b] + (1-\rho^*)\left(P_{ae}E[X_a] + (1 - P_{ae})E[X_u^*]\right), \tag{14}$$

where ρ^* equal to $\lambda E[X^*]$ stands for the probability of system being not empty and P_{ae} represents the average probability that the server is available when the system is empty. From [6, Eq. (23)], P_{ae} can be found equal to:

$$P_{ae} = 1 - \frac{(1 - \tilde{F_Y}(\lambda))(1 - \tilde{F_R}(\lambda))}{\lambda E[Y](1 - \tilde{F_Y}(\lambda)\tilde{F_R}(\lambda))}, \tag{15}$$

where $\tilde{F}_Z(.)$, for an arbitrary distribution function F_Z, can be defined as [6, P.753]:

$$\tilde{F}_Z(\theta) = \int_0^\infty e^{-\theta u} dF_Z(u). \tag{16}$$

For $\theta = \lambda$ and $Z = Y$ or R, this function gives the probability that the length of the operating and recovery periods is less than an inter-arrival time. As expected, when λ is small so that the inter-arrival times are long compared to Y and R, several operating and recovery periods can be observed during an inter-arrival period. Therefore, for small λ, $E[R]$ and $E[Y]$, we can write:

$$P_{ae} \approx P_a = \frac{E[Y]}{E[Y] + E[R]} \text{ , (For small } \lambda). \tag{17}$$

In this case, since P_{ae} is a function of $E[Y]$ and $E[R]$ and Y and R are assumed stationary processes, we can assume P_{ae} independent of Y. Also in general and for the regular distributions, P_{ae} is a function of only the moments of Y and R. Therefore, we assume for simplicity that there is no correlation between P_{ae} and Y and R. Note that P_{ae} is a

conditional probability, conditioned on the fact that the system is empty.

We also use the notation of X_e^* for the completion time of a packet which arrived to an empty system (cases 'a' and 'u' together). Its expected value can be given by:

$$E[X_e^*] = P_{ae}E[X_a] + (1 - P_{ae})E[X_u^*]. \tag{18}$$

The second moment of the general completion time can be found equal to:

$$E[X^{2*}] = \rho E[X_b^2] + (1 - \rho)E[X_e^{2*}]. \tag{19}$$

Now to derive the waiting time, we use the same approach as an M/G/1 queue. When a packet arrives, it waits for the remaining completion time of the packet in service (if any) and then the completion time of all packets in the queue. The point is that for the packets which are in the queue, the completion time is always distributed with X_b (they are queued, so they have not arrived to an empty system). However for the packet which is in service, the general completion time should be used because no knowledge is available to know whether this packet has been of case 'a', 'b' or 'u'. Thus, we can write:

$$E[W^*] = \frac{\lambda E[X^{2*}]}{2(1 - \lambda E[X_b])}. \tag{20}$$

We can see that the average remaining completion come for a packet in service (similar to the remaining service time in an M/G/1 queue) can be written equal to:

$$E[J_C] = (\rho^*)\frac{E[X^{2*}]}{2E[X^*]} = \lambda\frac{E[X^{2*}]}{2}. \tag{21}$$

E. Second Model of Case 'u'

We can all the parameters above for the the second model of case 'u'. $E[X]$ and $E[X^2]$ can be found similarly by replacing X_u^* with X_u. As discussed, J_C and $E[T] = E[W] + E[X] = E[W^*] + E[X^*]$, remain the same. Therefore, $E[W]$ can be found as follows:

$$E[W] = E[W^*] + E[X^*] - E[X]. \tag{22}$$

F. Busy Periods

Using the same idea of calculating the distribution of the busy periods in an M/G/1 queue without interruption [8], we can find the busy periods of our queue with interruption. We know that the first completion time in a busy period is an instance of X_e. However, for other busy periods which are initiated during X_e (the idea of serving in an LCFS manner [8]), the busy period is started with an instance of X_b because the packets enter a non-empty system. Therefore, we can find the LST of the busy periods equal to:

$$\hat{B}(s) = \hat{X}_e(s + \lambda - \lambda\hat{B}_b(s)), \tag{23}$$

where $\hat{B}_b(s)$ is the LST of the busy periods which are initiated during X_e thus with an instance of X_b. Therefore, $\hat{B}_b(s)$ itself can be found from the following equation:

$$\hat{B}_b(s) = \hat{X}_b(s + \lambda - \lambda\hat{B}_b(s)). \tag{24}$$

From the equation above, we can find the first and the second moments of the $B_b(t)$ and from there, the first and the second moments of the general busy periods, equal to:

$$E[B] = \frac{E[X_e]}{1 - \lambda E[X_b]}, \tag{25}$$

and

$$E[B^2] = \lambda E[B_b^2]E[X_e] + (1 + \lambda E[B_b])^2 E[X_e^2]. \tag{26}$$

The moments of the busy periods which start with an instance of X_b, $E[B_b]$ and $E[B_b^2]$, can be found from solving Eq. (24) equal to:

$$E[B_b] = \frac{E[X_b]}{1 - \lambda E[X_b]} \text{ and } E[B_b^2] = \frac{E[X_b^2]}{(1 - \lambda E[X_b])^3}. \tag{27}$$

Similarly, we can define the busy periods which are initiated with any initial delay. Note that similar to remaining completion time, the busy periods are the same in both models because even in the second model, the arrival of the packet should be considered as the beginning of the busy period.

G. Approximation for the Distribution of Y_{1a} and Y_{1b}

To find Y_{1a}, as discussed in [6], we should consider two possibilities. When the last packet before our packet in consideration, which has just arrived, has left the system in the same availability period, and the case that the last packet has left sooner, so during this availability period, the system has been empty. In the former case, we can see that the remaining time of the availability period after the departure of the last packet before us is distributed as Y_{1b}. The time between its departure and the next arrival (of our packet in consideration) is Exponential with parameter λ. Therefore, the remaining time of the availability period can be written as:

$$\begin{aligned} Y_{1a}|(&\text{Last departure in the same availability period}) \\ &= Y_{1b} - A|(Y_{1b} > A). \end{aligned} \tag{28}$$

In the latter case, we can simply replace Y_{1b} with Y. We have:

$$\begin{aligned} Y_{1a}|(&\text{Last departure in previous availability periods}) \\ &= Y - A|(Y > A). \end{aligned} \tag{29}$$

Finding the probability of occurrence of each case is cumbersome and, as discussed in [6], needs some simplification assumptions for Y_{1b}. As the distribution of Y_{1a} and Y_{1b} is complicated, we can follow two approaches. The first one is finding lower and upper bounds, as discussed in [6, Eq. (5)], for Y_{1b} and consequently for the general completion time.

The next approach is proposing an approximation for Y_{1a} by assuming that Y_{1a} may be started uniformly during an operating period which is sometimes called *random modification* of Y [7] or *equilibrium excess distribution* [6]. We can write the distribution of Y_{1a} based on Y equal to:

$$f_{Y_{1a}}(t) = \frac{1 - F_Y(t)}{E[Y]}. \tag{30}$$

The n-th moments of Y_{1a} based on the moments of Y can be written as [12]:

$$E[Y_{1a}^n] = \frac{E[Y^{n+1}]}{(n+1)E[Y]}. \tag{31}$$

Note that this is an approximation for Y_{1a}; thus, it is expected that we observe some difference between the simulation and analytical results. By this assumption, we approximate that X_b has almost the same distribution as X_a. That is, $E[X_b] \approx E[X_a]$ and $E[X_b^2] \approx E[X_a^2]$. Note that by this assumption for Y_{1b}, the probability of the occurrence of each one of the possibilities for Y_{1a} (Eqs. (28) and (29)) can be found equal to [6]:

$$\begin{aligned} &Pr(\text{Arrival in the same availability period}) \\ &\qquad = 1 - \int_0^\infty e^{-\lambda t} f_{Y_{1b}}(t)dt. \end{aligned} \tag{32}$$

This approximation helps us to have an *equilibrium renewal process* for cases 'a' and 'b' [12] which is much easier to calculate. The LST of the moments of the number of renewals with this assumption can be written as:

$$\mathcal{L}[m_a(t)] = \frac{1}{E[Y]S^2} \Longrightarrow m_a(t) = \frac{t}{E[Y]}, \tag{33}$$

and

$$\mathcal{L}[m_a^2(t)] = \frac{1 + Y(S)}{S^2 E[Y]\,(1 - Y(S))}. \tag{34}$$

The unconditional moments can then be found similar to Eqs. (6) and (7), but with the new distributions for $m_a(t) \approx m_b(t) = \frac{t}{E[Y]}|(T = t)$ and for $m_a^2(t) \approx m_b^2(t)$.

H. Exponential Availability Periods

As in the paper it is mostly assumed that availability periods are distributed with an Exponential distribution, we discuss this case separately here. Y is Exponentially distributed with the parameter α (i.e., $F_Y = F_{Y_{1a}} = 1 - e^{-\alpha t}$) thus we have $Y_{1a} = Y_{1b} = Y$ and $K_a = K_u = K_b$ because the remaining time in an availability period is still the same exponential distribution.

Using Eq. (3), $m(t)$ can be found equal to αt. From Eq. (5), the second moment of the number of renewals, $m^2(t)$, is equal to $m^2(t) = \alpha^2 t^2 + \alpha t$. These results verify our relations since the number of renewals for exponentially distributed availability periods is a Poisson process with the average αt. Therefore, we have:

$$\begin{aligned} E[X_a] &= E[X_b] = E[X_u] \\ &= E[T] + \alpha E[T]E[R] = E[T](1 + \alpha E[R]), \end{aligned} \tag{35}$$

$$E[TK_a] = \int_0^\infty \alpha t^2 f_T(t) = \alpha E[T^2], \tag{36}$$

$$E[X_b^2] = E[T^2](1 + \alpha E[T])^2 + \alpha E[T]E[R^2], \tag{37}$$

$$E[X_u^{2*}] = E[X_a^2] + E[R_r^2] + 2E[X_a]E[R_r]. \tag{38}$$

An interesting point when the availability periods are Exponentially distributed is the fact that considering the second model assumed for case 'u', the completion time of all packets is the same (equal to X_b). So, to find the waiting time, we can model our queue as a queue with vacation. That is, the recovery period in the arrival point of the packet which initiates a busy period is modeled as vacation of the server to use the available results in the literature for M/G/1 queues with vacations (queues with initial setup time) [8]. The initial setup time for a packet which initiates the busy period is R_r with the probability $(1 - P_{re})$ and zero otherwise. Thus, we can find the moments of the initial setup time based on the moments of Y and R. That is, $E[S] = (1 - P_{ae})E[R_r]$ and $E[S^2] = (1 - P_{ae})E[R_r^2]$. From [8, Eq. 2.44a], we have:

$$\begin{aligned} E[W] &= \frac{\lambda E[X_b^2]}{2(1 - \lambda E[X_b])} + \frac{2E[S] + \lambda E[S^2]}{2(1 + \lambda E[S])} \\ &= \frac{\lambda E[X_b^2]}{2(1 - \lambda E[X_b])} + \frac{E[R^2]}{2(E[Y] + E[R])}, \end{aligned} \tag{39}$$

and

$$E[D] = E[X_b] + E[W]. \tag{40}$$

The steady-state probability of system being empty, P_0, can be given by:

$$P_0 = \frac{E[I]}{E[I] + E[B_s]} = \frac{1 - \lambda E[X_b]}{1 + \lambda E[S]}, \tag{41}$$

where $E[I] = \frac{1}{\lambda}$ is the average of Idle periods (no packet in the system), and $E[B_s]$ is the average of busy periods initiated by $S + X_b$ which can be found from Section II-F.

III. Priority Queueing

The results found so far can be used to tackle priority queueing disciplines over this general queueing model with N_p classes of traffic. We define and apply four different preemption schemes: classic schemes of *non-preemptive* and *preemptive-resume* [14], an exceptional non-preemptive and a new one which is called *preemption in case of failure*. A non-preemptive scheme permits a low priority (LP) packet to be totally transmitted before serving a new arrived high priority (HP) packet. This implies that even if an LP packet arrives in an empty system during the recovery, incoming HP packet in the same recovery period should also wait for the termination of the LP service. While in the exceptional non-preemptive scheme, the HP packet which is arrived in the same arrival recovery period starts the service. But, If LP starts the real service, the service can not be preempted. A preemptive discipline preempts the service from the LP packets in any time as soon as an HP packet arrives. The new scheme which, to the best of our knowledge, is a novel scheme and very well applicable to cognitive radio networks, does not preempt the service during the real transmission of the LP packet. But, if a failure occurs which necessitates a recovery, after the recovery, the HP packets (if any) will be served first. When the HP queue is empty again, the transmission of the LP packet from the same state is resumed.

For simplicity of presentation, we will use parameters with an index 1 for high priority (HP) packets and with an index 2 for low priority (LP) ones. We focus on two classes however it can be, sometimes, extended to multiple classes of traffic. Furthermore, in the rest of the paper, when notations λ, ρ

and A are used without any index, they represent combined parameters. That is:

$$\lambda = \sum_{j=1}^{N_p} \lambda_j, \quad (42)$$

$$\rho = \sum_{j=1}^{N_p} \rho_j = \sum_{j=1}^{N_p} \lambda_j E[X_b^j], \quad (43)$$

and

$$A = \min\{A_1, ..., A_{N_p}\} \rightarrow A \sim EXP(\lambda). \quad (44)$$

We also use the notation of $P_{ae}(\lambda)$ and $R_r(A)$ to highlight that P_{ae} and R_r in Eqs. (15) and (13), respectively, should be calculated with combined λ and A.

A. Non-preemptive Priority

In a non-preemptive scheme, when any packet from any class goes to service, its service will not be interrupted by other packets. This implies that the completion time of any packet of any class will be the same as the original M/G/1 queue with interruptions, as already discussed. However, even though the completion time of all cases is the same as the original queue, the average completion time for the first model (with '*') is different because the probability of occurrence of each case ('a', 'b' and 'u') is different. That is, in a queue with multiple classes of traffic and when the service can not be preempted from the packet in service, the probability of arrival in an empty system is less. Thus, X_b plays a more determining role and consequently the average completion time in such a queue is less, compared to a queue with a single classes of traffic. However, the waiting time naturally increases.

1) First Approach: If we return to the original queue and replace the probability of system being empty from $\lambda E[X^*]$ to $\sum_{j=1}^{N_p} \lambda_j E[X_j^*]$ in Eq. (14) for all classes of traffic, we will have N_p equations and N_p unknowns which enables us to find the new average completion time of the packets in this system. To find P_{ae}, λ in Eq. 15 is replaced with $\sum \lambda_j$ for all classes of traffic. Consequently, the second moment can be found. Then, similar to an M/G/1 queue [14] we have:

$$E[W_i^*] = \frac{E[J^*]}{(1 - \sum_{j=1}^{i} \rho_j)(1 - \sum_{j=1}^{i-1} \rho_j)}. \quad (45)$$

We used the fact that $C_{YR}(T) = E[X_b]$ for all queued packets. J^* stands for the remaining completion time of the packet in service, which is the same for all priority classes. It can be given by:

$$E[J^*] = \sum_{j=1}^{N_p} \frac{\lambda_j}{2} E[X_j{}^{*2}], \quad (46)$$

where N_p is the number of priority classes. This implies that for the packet in service, no knowledge is available whether it has been a packet of case 'a', 'b' or 'u', so a the general completion time is used. However, the denominator represents the completion time of the queued packets which is X_b^j for class j.

When Y and arrival processes are all exponentially distributed, which is of our interest in this paper, we know that $X_b = X_a = X_u$. Thus, using the idea of Eq. (39), a closed-form relation can be given for the waiting time of the packets. For this aim, the queue can be modeled as a queue with an exceptional completion time for the first packet which initiates a busy period. As saw for the original queue, this exceptional service is represented by X_e^* (Eq. (19)). However here, X_e^* should be updated by the combined probability of arrival during Y or R. We define:

$$X_e^{i*} = P_{ae}(\lambda)(X_b^i) + (1 - P_{ae}(\lambda))(X_b^i + R_r(A)), \quad (47)$$

$$X_e = \sum_{j=1}^{N_p} \frac{\lambda_i}{\lambda} X_e^{i*}. \quad (48)$$

First two moments of X_e can be found easily. Then, we can write [8]:

$$E[T_i] = \frac{(1-\rho_b)(E[X_e^{i*}] + E[X_b^i]) + \lambda E[X_e^{i*}]E[X_b^i]}{1 + \lambda E[X_e] - \rho} + \frac{\lambda[(1-\rho)E[X_e^2] + E[X_e](\sum_{j=1}^{N_p} \lambda_j E[(X_b^j)^2])]}{2(1 + \lambda E[X_e] - \rho)(1 - \sum_{j=1}^{i} \rho_j)(1 - \sum_{j=1}^{i-1} \rho_j)}. \quad (49)$$

The first term represents the average completion time and the second term, the average waiting time. Note that the equation above is a closed-form relation of the general case (provided in Eq. (45)), for exponential availability periods.

Similar to Eq. (41), the steady-state probability of system being empty, P_0, in this queue can be given by:

$$P_0 = \frac{1-\rho}{1 - \rho + \lambda E[X_e]}. \quad (50)$$

2) Second Approach: Another idea is to use available results in the literature for the mixed priority queueing disciplines. As interruptions have a preemptive behavior, we can model the interruptions as the highest priority type of traffic whose inter-arrival times are distributed with a random variable Y and whose *busy periods* are distributed with a random variable R. From these two, we can have an estimate for the service time of these virtual highest priority packets whose activity models the interruptions. Closed-form relations can not be derived in general since for any distribution of R and Y, a different formula for the busy periods and consequently for the service time exists. If Y is Exponentially distributed with parameter α, we can assume an original M/G/1 queue for the interruptions. Therefore, from the following equation, which is the distribution of the busy periods of a regular M/G/1 queue [8], we can find the first two moments of the service time of the virtual packets which form the interruptions:

$$\hat{R}(s) = \hat{T}_v(s + \alpha - \alpha\hat{R}(s)), \quad (51)$$

$$E[T_v] = \frac{E[R]}{1 + \alpha E[R]} \;,\; E[T_v^2] = E[R^2](1 - \alpha E[T_v])^3. \quad (52)$$

where T_v stands for the service time of the virtual packets who form the interruptions. Then, the queue can be modeled as a priority queue where the highest class of traffic (virtual) behaves preemptively, but other classes behave non-preemptively. From ([8]), the extensions of the P-K formula can be used

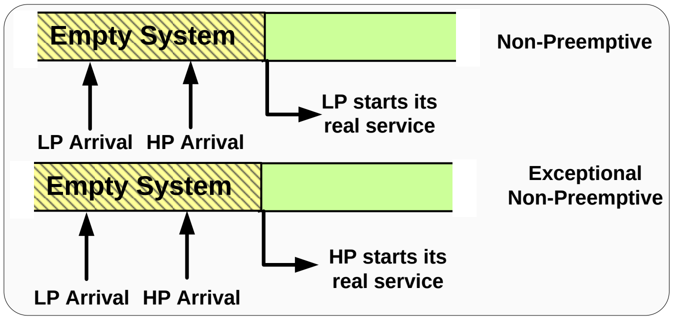

Fig. 5. The difference between Non-preemptive and Exceptional Non-Preemptive schemes.

to find the moments of the waiting time of other classes of traffic. For the class i, it can be given by:

$$E[W_i] = \frac{E[J]}{(1-\alpha E[T_v]-...-\lambda_{i-1}E[T_{i-1}])(1-\alpha E[T_v]-...-\lambda_i E[T_{i-1}])}, \tag{53}$$

where $E[J]$ is equal to:

$$E[J] = \frac{1}{2}\alpha E[T_v^2] + \sum_{j=1}^{N_p} \frac{1}{2}\lambda_j (E[T_j^2]). \tag{54}$$

The system time can be written as:

$$E[D_i] = \frac{E[T_i]}{1-\alpha E[T_v]} + E[W_i]. \tag{55}$$

Note that this approach provides an approximation because we know that the real distribution of the busy periods in an M/G/1 queue is a complicated function built on the Bessel function [8]. For other distributions of availability periods, if we are able to find the probability generating function (PGF) of the number of arrivals, we can apply the same approach to find the moments of the service time of the virtual packets which form the interruptions.

B. Exceptional Non-Preemptive

If we remember two different models which discussed in the original queue (with and without '*'), we can discuss a new non-preemptive scheme with a small difference that if before the start of the real service time of an LP packet, an HP packet arrives, the HP packet is served first. As illustrated in Figure 5, this case happens only when an LP packet arrives into an empty system and during the recovery period (case 'u' or the original queue), and an HP packet also arrives in the same recovery period. In other words, this the a non-preemptive scheme when the completion time is defined without '*' (start of the real service).

Similar to the previous discipline, this queue can be modeled as a queue with initial setup time with the difference that during the setup time, the service is preemptive [8]. As the arrival process is Poisson for both classes of traffic, the setup time is the same for both classes and similar to the previous discipline. The service time of different classes of packet can be given by:

$$E[W_i] = \frac{\sum_{j=1}^{N_p}\lambda_j E[(X_b^j)^2]}{2(1-\sum_{j=1}^{i}\rho_j)(1-\sum_{j=1}^{i-1}\rho_j)} + \frac{(1-\rho)(\lambda E[S^2]+2E[S])}{(2(1+\lambda E[S]-\rho))(1-\sum_{j=1}^{i}\rho_j)(1-\sum_{j=1}^{i-1}\rho_j)}, \tag{56}$$

$$E[D_i] = E[W_i] + E[X_b^i]. \tag{57}$$

The probability of system being empty is equal to:

$$P_0 = \frac{1-\rho}{1+\lambda E[S]}. \tag{58}$$

C. Preemptive-resume Priority

In this scheme, the highest class is not affected by other classes of traffic. So, the completion time and the waiting time of the highest priority packets can be found from the original M/G/1 queue with interruptions. Let us find the completion time and the waiting time of other classes.

1) First Approach: An approach is to use available results in the literature for the preemptive-resume schemes with multiple classes of traffic, similar to the one that we used for the previous scheme. As interruptions have a preemptive behavior and we are discussing the preemptive schemes, we can model the queue as a preemptive-resume queue with N_p+1 classes of traffic. Then the following famous extensions of the P-K formula for preemptive-resume schemes ([8]) can be used to find the moments of the waiting time of other classes of traffic:

$$E[W_i] = \frac{E[J_i]}{(1-\alpha E[T_v]-...-\lambda_{i-1}E[T_{i-1}])(1-\alpha E[T_v]-...-\lambda_i E[T_{i-1}])}, \tag{59}$$

where $E[J_i]$ can be given by:

$$E[J_i] = \frac{1}{2}\alpha E[T_v^2] + \sum_{j=1}^{i} \frac{1}{2}\lambda_j (E[T_j^2]), \tag{60}$$

and the moments of T_v can be found from Eq. (52).

2) Second Approach: Another approach, which is provided here because it will also be used for the third priority scheme, is to find the distribution of availability and unavailability periods from the perspective of the LP packets. Let us call them respectively $Y2$ and $R2$. Here, unavailability implies both due to the activity of HP users and due to interruptions. As illustrated in Figure 6.a, it is clear that $Y2$ is the minimum of two random variables A_1 and Y which means between the interruptions and the HP packets, any one which arrives sooner initiates an unavailability period for LP packets. A_1 stands for the inter-arrival time of the HP packets. Therefore,

$$Y2 = \min(Y, A_1) \rightarrow 1-F_{Y2}(t) = (1-F_Y(t))(1-F_{A_1}(t)). \tag{61}$$

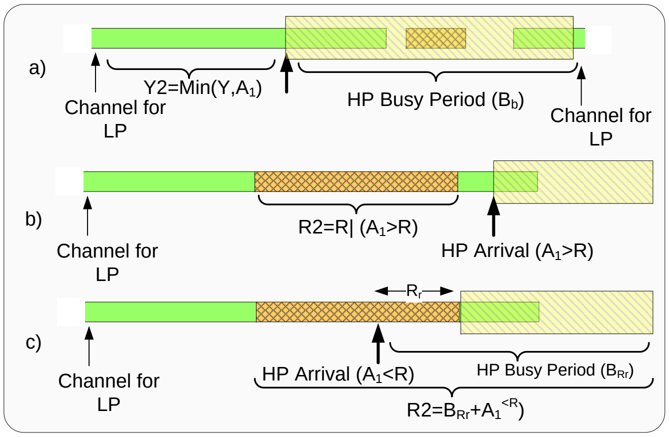

Fig. 6. Availability and unavailability periods ($Y2$ and $R2$) from the perspective of LP packets.

When Y is Exponentially distributed with parameter α, the distribution of $Y2$ can be given by:

$$F_{Y2}(t) = 1 - e^{-(\lambda_1+\alpha)t}. \tag{62}$$

To calculate $R2$, we should take into account which one of two events discussed above has occurred sooner. As can be seen in Figure 6.a, if the HP packet arrives sooner, the length of $R2$ is equal to one busy period of HP packets which is distributed with B_b. However, if an interruption arrives sooner, again two different cases may happen which discussed in sections b and c of Figure 6. During R, no HP packet arrives, so the length of $R2$ is equal to R (conditional), or at least an HP packet arrives. In the latter case, the length of the unavailability period from the perspective of the LP packets will be the interruption period which is extended with an HP busy period. We need the remaining time of the interruption period after the arrival of the HP packet. Recalled from Eq. (13), this time is distributed as $R_r = R - A_1|(R > A_1)$. After this time, an HP busy period is started with a duration equal to B_{R_r}. B_{R_r} can be found from Eq. (23), replacing X_b by $X_b + R_r$. We have:

$$R2 = \begin{cases} B_b & Pr(A_1 < Y), \\ \begin{cases} R|(R \le A_1) & Pr(Y \le A_1 \& R \le A_1)), \\ A_1|(R > A_1) + B_{R_r} & Pr(Y \le A_1 \& R > A_1). \end{cases} \end{cases} \tag{63}$$

From now on and for simplicity, we use the notation $Z1^{<Z2}$ for $Z1|(Z1 < Z2)$ for any two random variables $Z1$ and $Z2$. The probability of an HP arrival during R (i.e., $A_1 < R$) can be given by:

$$\begin{aligned} P_{a1iR}|(R = r) &= 1 - e^{-\lambda_1 r} \rightarrow \\ P_{a1iR} &= \int_0^\infty (1 - e^{-\lambda_1 r}) dF_R(r). \end{aligned} \tag{64}$$

For Exponential Y, we can write:

$$Pr(Y \le A_1) = \frac{\alpha}{\alpha + \lambda_1}, \tag{65}$$

and then

$$\begin{aligned} E[R2] = \frac{\lambda_1}{\alpha+\lambda_1} E[B_b] + \frac{\alpha}{\alpha+\lambda_1} \Bigg[(1 - P_{a1iR}) \frac{-d/d\lambda_1 \hat{R}(\lambda_1)}{\hat{R}(\lambda_1)} \\ + P_{a1iR}(E[A_1^{<R}] + E[B_{R_r}]) \Bigg]. \end{aligned} \tag{66}$$

Using Eq. 92 for $E[R_r]$ and Eq. 25 for busy periods, we have:

$$E[B_{R_r}] = \frac{E[X_b^1] + E[R_r]}{1 - \lambda_1 E[X_b^1]} = \frac{E[X_b^1] + \frac{E[R]}{1-\hat{R}(\lambda_1)} - \frac{1}{\lambda_1}}{1 - \lambda_1 E[X_b^1]}. \tag{67}$$

The second moment of $R2$ can be found similarly, as the second moment of the busy periods can be found from Eq. (26) and the second moment of the conditional A_1 and R can be derived from Eq. (91) in Appendix. It should be taken into account that B_{R_r} and $A_1^{<R}$ are correlated, so $E[B_{R_r} A_1^{<R}]$ should be calculated separately, using, for instance, the same approach as in Eq. (8).

From the equations above, one can find the moments of the availability and unavailability periods from the perspective of LP packets ($Y2$ and $R2$). Then, we return to the original M/G/1 queue with interruptions and replace Y and R in Eq. (39) with $Y2$ and $R2$ respectively, to find the performance metrics of the LP packets.

D. *Preemption in Case of Failure*

In this priority queueing model, the type-1 packets can not preempt the service from a low priority packet when the latter is in transmission. However, the service can be preempted after an interruption. As expected, for both classes of traffic, the performance metrics are between two previous priority schemes. That is, for instance for the HP class, the average system time is more than the preemptive scheme but less than the non-preemptive scheme. Moreover, the completion time of the packets with the highest priority is not affected, so it can be still found from the original queue. However, the waiting time is affected. Other parameters which are the completion time of the LP packets and the waiting time of both HP and LP packets are discussed.

1) Completion Time of the LP Packets: We want to find the completion time of the LP packets. The same idea used for the preemptive scheme can be employed with the difference that here, the HP packet can not preempt the channel before the end of the cycle. However, the difficulty is that the length of the new interruption periods from the perspective of LP packets, $R2$,depends on the remaining service time of the LP packets which naturally is not the same during the completion time of an LP packet. Unless, we assume a memory-less service time which is a special case and will be discussed later. The problem can not be therefore modeled as a renewal process because the instances of $R2$ are not identical. Thus, we provide some approximations. The approximations are discussed for two cases: when the availability periods are much larger than the service time of type-2 packets ($Y >> T_2$), called *large scenarios*, and when it is smaller ($Y < T_2$), called *small scenarios*.

Using the first assumption, it can be assumed that the service of a type-2 packet is finished in maximum two type-2

operating periods. The selection of two is a trade-off between accuracy and complexity. That is, if it is assumed that the service can be finished in only one operating period, the bound will be very loose, and if three type-2 operating periods are considered, the complexity increases.

Assuming two type-2 operating periods is equivalent to experiencing at most one $R2$ interruption during the completion time. The duration of the interruption depends on the arrival of an HP packet, as already discussed. We can thus write the completion time of type-2 packets as:

$$E[X_2] \approx \begin{cases} E[T_2] & Y \geq T_2 \\ E[T_2] + E[R2] & Y < T_2. \end{cases} \tag{68}$$

where $R2$ stands for the length of the interruption period from the perspective of type-2 users. It can be given by:

$$R2 = \begin{cases} R^{<A_1} & \text{No HP arrival,} \\ B_{C_r} - (Y^{<T_2} - A_1|(A_1 < Y^{<T_2})) & \text{HP arr. in } Y^{<T_2}, \\ A_1^{<R} + B_{R_r} & \text{HP arr. in } R. \end{cases} \tag{69}$$

where C_r stands for the remaining time of the cycle after the arrival of an HP packet, and B_{C_r} represents the busy period which is initiated with $C_r + X_b$. However, if we return to Figure 6.a, we can see that the remaining time of $Y^{<T_2}$, which is represented by $Y^{<T_2} - A_1$, should be excluded from $R2$. In the third case, the HP arrival occurs in R. So, the busy period of HP packets is started with $R_r + X_b$ and the length of the total interruption is A_1 in addition to HP busy period. Thus, $E[R_2]$ can be given by:

$$\begin{aligned} E[R2] = {} & (1 - P_{a1ic^<}) \frac{-d/d\lambda_1 \hat{R}(\lambda_1)}{\hat{R}(\lambda_1)} \\ & + (P_{a1ic}) \Bigg[P_{ae} \Big(\frac{E[R] + E[Y^{<T_2} - A_1|(< Y^{<T_2})] + E[X_b^1]}{(1 - \lambda_1 E[X_b^1])} \\ & - E[Y^{<T_2} - A_1|(A_1 < Y^{<T_2})] \Big) \\ & + (1 - P_{ae})(E[A_1^{<R}] + \frac{E[X_b^1] + E[R_r]}{(1 - \lambda_1 E[X_b^1])}) \Bigg], \end{aligned} \tag{70}$$

where $P_{a1ic^<}$ is the probability of an arrival in $C^< = Y^{<T_2} + R$, and P_{ae} is calculated for HP packets.

The second moment of $R2$ can be found similarly using the second moment of the busy periods and the relations provided in Appendix. However, again the correlation of random variables B_{R_r} and $A_1^{<R}$ should be taken into account.

For the other approximation (small scenarios), as Y is supposed to be short, we assume that the duration of the HP busy periods is independent of the activity of LP packets. Therefore, the interruption periods from the perspective of LP packets can be assumed to have a similar distribution and thus, a renewal process can be considered. This implies that Eq. (69) is still valid however we eliminate the condition that $Y < T$. When a type-2 packet starts its service, as illustrated in Figure 7, it can hold the channel (available or unavailable) for a cycle and if there is an HP arrival during the cycle, it releases the channel to HP packets at the end of the cycle. So, the probability of releasing the channel to HP traffic in the next cycle is the probability of at least one HP arrival with the rate λ_1 during the current cycle. We can find this probability equal to:

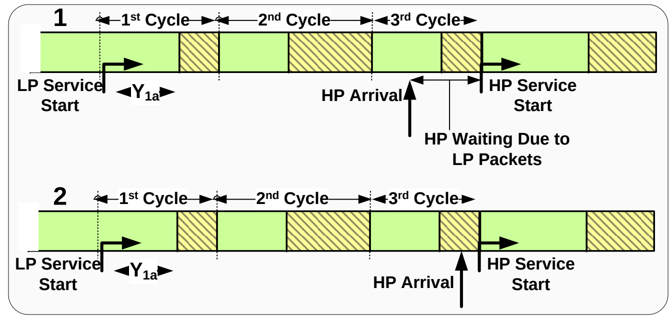

Fig. 7. Cycles and holding periods for the LP packets in the discipline of preemptive in case of failure (FP). The LP packet can hold the channel for three cycles; however in a preemptive scheme (Pr), it can keep the channel for two cycles and releases the channel in the middle of the third cycle (HP arrival).

$$\begin{aligned} P_{a1ic} &= Pr(\text{HP arrival in } C) \\ &= \int_0^\infty (1 - e^{-\lambda_1 c}) f_C(c) dc. \end{aligned} \tag{71}$$

The interruption periods can be discussed similar to the preemptive scenario, as it is again assumed a renewal process. We have:

$$R2 = \begin{cases} R|(C < A_1) & \text{No HP arrival in } C, \\ A_1|(C > A_1) + B_{C_r} & \text{HP arrival in } C. \end{cases} \tag{72}$$

where C_r stands for the remaining time of the cycle after the arrival of an HP packet. Due to the independence of Y and R, $R|(C < A_1) = R|(R < A_1)$, which was found in Section III-C2. Similar to previous assumption, C_r itself is discussed separately for two cases: when the arrival has occurred in Y and when it has occurred in R, during the unavailability period.

When the HP packet arrives in Y, the length of the remaining time of the cycle can be given by the original cycle length, $C_r = C = Y + R$. Because, the remaining time of the availability period is the same as Y (memoryless property) and there is no condition for R. However, when the arrival occurs in R, we return to part (c) of Figure 6. That is, $C_r = R_r$. It can thus be written:

$$\begin{aligned} E[R2] = {} & (1 - P_{a1ic}) \frac{-d/d\lambda_1 \hat{R}(\lambda_1)}{\hat{R}(\lambda_1)} \\ & + (P_{a1ic}) \Bigg[P_{ae} \Big(\frac{E[R] + E[Y]}{(1 - \lambda_1 E[X_b^1])} - E[Y] \Big) \\ & + (1 - P_{ae})(E[A_1^{<R}] + \frac{E[X_b^1] + E[R_r]}{(1 - \lambda_1 E[X_b^1])}) \Bigg]. \end{aligned} \tag{73}$$

The second moment can be found similarly. The moments of $R2$ then can replace R in Eqs. 35 and 37 to find the moments of the completion time of LP packets ($E[X_2]$ and $E[X_2^2]$).

2) HP Waiting Time: We can see that in this scheme, only HP packets of case 'a' (HP arrival during Y to the system which is empty of HP packets) are affected. For these packets, whose waiting time was zero in the previous case, the new waiting time will be a function of the LP queue. If the LP

queue is empty, the waiting time is naturally zero. Otherwise, the waiting time is the minimum of the remaining time of the arrival cycle and the remaining time of the service of the LP packet in service.

The difficulty is the dependency of both types of traffic on each other. That is, while the waiting time of the HP packets are affected by the lower class, both the waiting time and the completion time of LP packets are affected by HP traffic. This obliges us to suffice to approximations and bounds.

For the exponentially distributed availability periods, which is the assumption in this part, we can see that the upper bound for the waiting time of HP packets is an M/G/1 queue with vacation (similar to the original queue) in Eq. (39)when S (initial setup time) is updated, as follows:

$$S = \begin{cases} R_r(\lambda_1) & (1 - P_{ae}(\lambda_1)), \\ Y^{<T_2} + R & Pr(Y \leq T_2)(1 - P_0)P_{ae}, \\ T_2^{<Y} & Pr(Y > T_2)(1 - P_0)P_{ae}, \\ 0 & P_0 P_{ae}, \end{cases} \tag{74}$$

where $P_{ae}(\lambda_1)$ and $R_r(\lambda_1)$ take only HP packets into account (not the total arrival rate), and P_0 is the probability of system being empty of any type of packet. We can write:

$$\begin{gathered} \frac{\lambda E[X_b^2]}{2(1 - \lambda E[X_b])} + \frac{E[R^2]}{2(E[Y] + E[R])} \\ \leq E[W_1] < \\ \frac{\lambda E[X_b^2]}{2(1 - \lambda E[X_b])} + \frac{2E[S] + \lambda_1 E[S^2]}{2(1 + \lambda E[S])}. \end{gathered} \tag{75}$$

This provides us an upper bound for the waiting time of the HP packets. As we will see later, by this bound we are assuming that HP packets should always wait for an LP packet (in case of arrival in Y) and the remaining LP service time is T_2. Note that the probability of system being empty of any packet, P_0, is the same for all priority queueing models discussed in this paper, regardless of the discipline. We have found this probability in Eq. (50).

3) LP Waiting Time: When the moments of $R2$ and consequently X_2 are known, a relation similar to Eq. (39) can be used to find the waiting time of LP packets (M/G/1 with vacation). To calculate the moments of vacation in this case, consider the arrival of an LP packet to a system which is empty of LP packets. If the arrival occurs in the time that an HP packet exists in the system, the vacation is the general busy period of HP packets, which is not the same as the original queue with single class of traffic, and difficult to derive. In case of arrival to a system which is empty of HP packets, two different cases should be discussed: An LP arrival in Y implies that the waiting time is zero. If it occurs in R, two cases may happen. No HP arrival occurs in the remaining part of R (called R_r), so the waiting time is R_r, and an HP arrival occurs, so the LP packet waits for the end of the HP busy periods. However, finding the moments of all cases discussed is difficult due to the dependency of HP and LP packets, and is not in the scope of this paper. We thus propose some bounds, as follows.

Considering the Conservation Law (CL) in a queue with multiple classes of traffic [15], we know that

$$\kappa = \lambda_1 E[T_1]E[W_1] + \lambda_2 E[T_2]E[W_2] \tag{76}$$

is constant and the same for all priority disciplines proposed in this paper, regardless of the queueing discipline. In other words, as we found the waiting time of LP and HP packets in previous priority disciplines, we have the value of the sum above. If we find the waiting time of one of priority classes for the discipline of preemption in case of failure, we can find easily the other one.

As an approximation for the first two moments of the completion time of LP packets has been found, the minimum waiting time of LP packets can be found by P-K relation, without considering the impact of vacations for the packets which enter an LP empty system. LP packets experience maximum waiting time in the preemptive-resume model. Thus, we can write:

$$\begin{gathered} \frac{\lambda_2 E[X_2^2]}{2(1 - \lambda_2 E[X_2])} + \frac{E[R^2]}{2(E[Y] + E[R])} \\ < E[W_2] < \\ E[W_2] \text{ of preemptive scenario.} \end{gathered} \tag{77}$$

The Conservation Law helps us also to find another upper bound for the waiting time of HP packets, discussed in Eq. (75). The point is that the maximum waiting of HP packets is when LP packets experience the minimum waiting time. Thus, by replacing the lower bound of $E[W2]$ in Eq. (77) into Eq. (76) for E$[W_2]$, we can find another upper bound for HP packets.

E. Exponentially Distributed Service Times

The interesting point about the exponentially distributed service time is the fact that the remaining service time of the packets in each instant of time is still Exponential. This helps us to find more accurate relations for the last proposed scheme in which, only approximations were proposed. Discussion in [14] indicate that the performance results with an Exponential service time can be considered as the worst case among scenarios with other common distributions for service time (e.g., Erlang or Deterministic).

As we saw, the queue model for the HP traffic is an M/G/1 queue with vacation where we found an upper bound for it in Eq. (75) and by using the Conservation Law. Exponential service time of LP packets enables us to find the moments of the vacation from the perspective of HP traffic. The vacation may also exist when the HP packet arrives in Y and should wait for end of the service of LP packets or an interruption. The HP waiting time can thus be given by Eq. (39) when S (initial setup time) is updated, as follows:

$$S = \begin{cases} R_r(\lambda_1) & (1 - P_{ae}(\lambda_1)), \\ Y^{<T_2} + R & Pr(Y \leq T_2)(P_{L|NH})P_{ae}, \\ T_2^{<Y} & Pr(Y > T_2)(P_{L|NH})P_{ae}, \\ 0 & Otherwise. \end{cases} \tag{78}$$

where $P_{ae}(\lambda_1)$ and $R_r(\lambda_1)$ take only HP packets into account (not the total arrival rate). $E[S]$ can be written equal to:

$$E[S] = (1 - P_{ae})E[R_r] + (P_{L|NH})P_{ae}\left[(\frac{1}{\gamma+\alpha} + \frac{\alpha}{\alpha+\gamma}E[R])\right]. \quad (79)$$

The unknown here is $P_{L|NH}$, which is the probability that the system is empty of HP packets, but there are LP packets in the system, in the point of arrival of an HP packet. The probability of system being empty of HP packets, P_0^1 can be found from the original queue, replacing $E[S]$ with the one that was calculated in (Eq. 79). Using these two, we can write:

$$P_{L|NH} = 1 - Pr(\text{System empty if no HP packet}) = 1 - \frac{P_0}{P_0^1}. \quad (80)$$

As P_0 is known (Eq. 50), we have two equations (Eq. 79 and 80) and two unknowns ($E[S]$ and $P_{L|NH}$), so they can be found.

For the LP packets, the waiting time can be found easily from the conservation law. To find the completion time, the same approach which was used as an approximation for small availability duration (Eq. 73) can be used here because the exponentially distributed service time of LP packets makes the process for LP packets a renewal process. In other words, the distribution of interruptions, from the LP packets point of view, is identical. The only change is that here Y should be replaced with $Y^{<T_2} = Y|(Y < T_2)$ because the completion time continues if the availability period is smaller than the remaining service time of the LP packet (T_2). We have:

$$P_{aiY<T} = \frac{\lambda_1}{\lambda_1 + \alpha + \gamma}, \quad (81)$$

$$P_{aiR} = (1 - P_{aiY<T})Pr(R < A_1), \quad (82)$$

$$\begin{aligned} P_{NaiC<} &= (1 - P_{aiY<T} - P_{aiR}) \\ &= \frac{\alpha+\gamma}{\lambda_1+\alpha+\gamma}(1 - Pr(R < A_1)), \end{aligned} \quad (83)$$

$$E[B_{C_r}] = \frac{E[R] + \frac{1}{\gamma+\alpha} + E[X_b^1]}{1 - \lambda_1 E[X_b^1]}, \quad (84)$$

$$E[B_{R_r}] = \frac{E[R_r] + E[X_b^1]}{1 - \lambda_1 E[X_b^1]}, \quad (85)$$

$$\begin{aligned} E[R] = P_{NaiC<}(E[R^{<A_1}]) + P_{aiY<T}(E[B_{C_r}] - E[Y_r^{<T}]) \\ + P_{aiR}(E[B_{R_r}] + E[A_1^{<R}]). \end{aligned} \quad (86)$$

Similarly, $E[R2^2]$ and consequently $E[X_2]$ and $E[X_2^2]$ can be found.

It is worth noting that for two classes of traffic, only the assumption of Exponential service time for LP packets is required to find the results above. The service time of HP packets can be general.

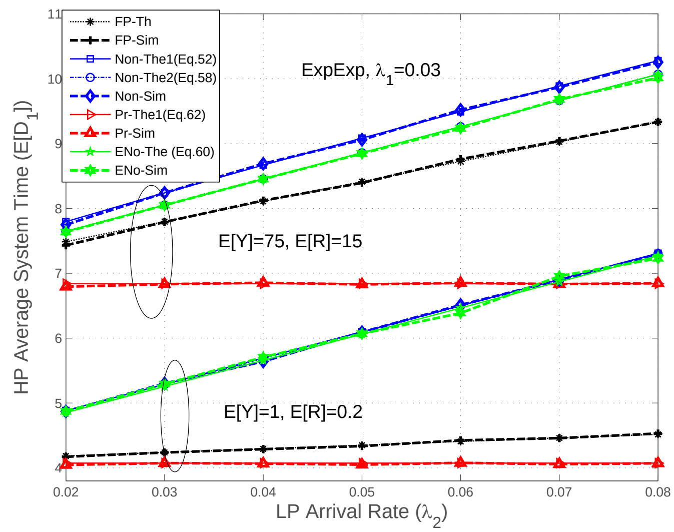

Fig. 8. System time of HP packets vs. LP arrival rate, for small and large scenarios.

IV. Simulation Results

In this section, simulation results for two different scenarios are provided: when interruption durations are distributed exponentially and when they are constant. In both scenarios, we assume two classes of traffic: high priority (HP) and low priority (LP), respectively known also as type-1 and type-2 packets. The service time (packet length) is also assumed either Exponential or constant. The scenarios are called 'ExpExp', 'ExpDet', 'DetExp' and 'DetDet' in the figures (respectively for T and R). Four disciplines discussed are called respectively 'Non' (non-preemptive), 'ENo' (exceptional non-preemptive), 'Pr' (preemptive) and 'FP' (preemption in case of failure) in the figures.

The duration of availability periods is selected in a way that it models realistic scenarios: when it is much larger than the service time of a single packet (called *large* scenario where $E[Y] = 75$ and $E[R] = 15$), which represents an almost static or lowly dynamic CR network, and when it is much smaller than the service time which represents a highly dynamic (opportunistic) CR network [16] (called *small* scenario where $E[Y] = 1$ and $E[R] = 0.2$). Their ratio, P_a, is selected the same in both scenarios.

We assumed that HP packets model signaling and control packets. Thus, their arrival rate, λ_1 is usually lower than LP packets and the packets are also smaller. HP arrival rate is assumed constant (unless mentioned) equal to 0.03 and we have $E[T_1] = 3$ and $E[T_2] = 5$. To model realistic scenarios, the unit of time can be assumed *milisecond* (ms) which represents, for instance, 378-Byte and 625-Byte packets over a link with a data rate equal to 1Mbps.

A. Exponential Recovery Time

Exponential recovery time is a realistic assumption when channel selection is done randomly, by sensing a list of channels one by one, and the user stops after finding the first

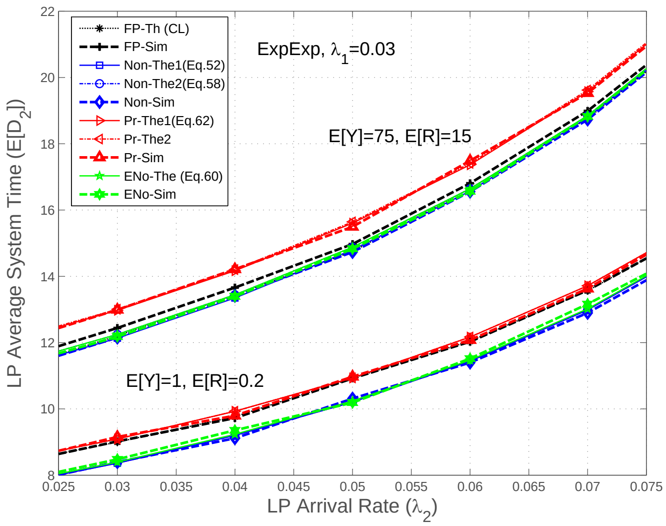

Fig. 9. System time of LP packets vs. LP arrival rate, for small and large scenarios.

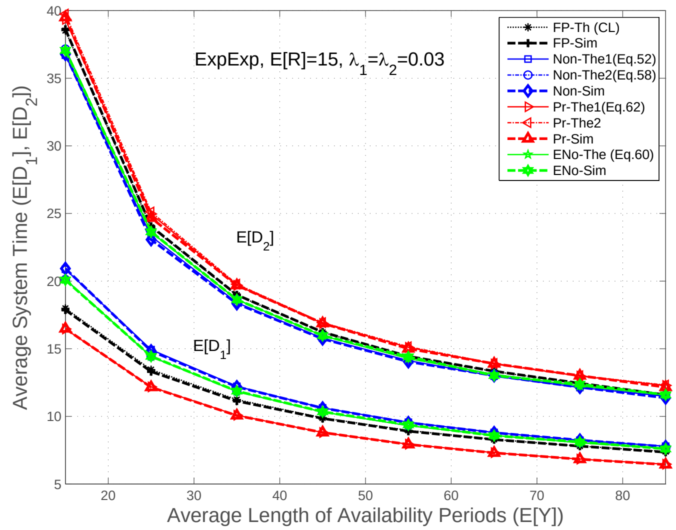

Fig. 10. System time of HP and LP packets vs. the variation of availability periods' duration (large scenarios).

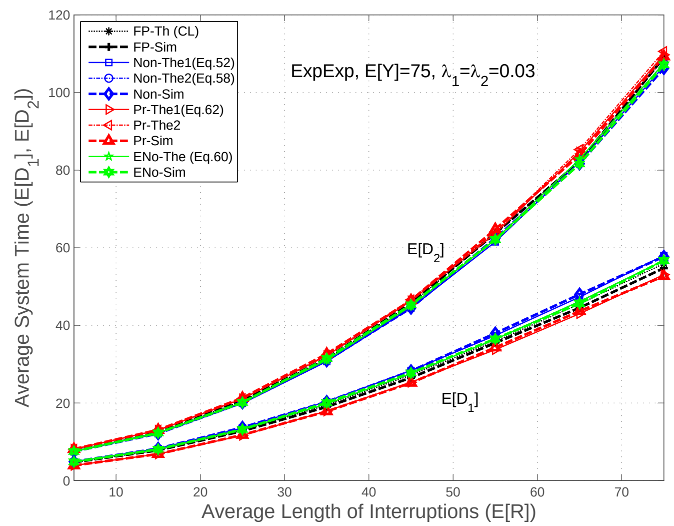

Fig. 11. System time of HP and LP packets vs. the variation of recovery periods' duration (large scenarios).

channel. The remaining time of the busy period after an arrival in R will be still Exponential. So, $R_r = R$.

As expected and can be verified in all figures from now on, the system time for any class of traffic using FP and ENo disciplines are between two other schemes, and ENo is generally close to non-preemptive (Non). As illustrated in Figures 8 and 9, when the packets length is Exponential (ExpExp), Eq. (78) can be used to find the waiting time of HP packets. Then, Conservation Law (CL) has been used to find the waiting time and consequently the completion time of LP packets for the FP scheme. Results are represented both for small and large availability periods. As expected, the system time of HP packets in preemptive scheme is independent of the arrival of LP packets, so it is constant. Moreover, we can see in both figures that with the same ratio of $E[Y]$ and $E[R]$, performance metrics are worse for all classes of traffic when availability and unavailability periods compared to packet length is large (large scenarios). The other point that can be verified is the performance difference of the exceptional non-preemptive (ENo) and preemptive in case of failure (FP) compared to other two schemes. In small scenarios, the service can be preempted very soon because interruptions are frequent. The FP discipline performance is thus close to the preemptive scheme. Toward large scenarios, FP performance gets farther from the Pr discipline and gets closer to non-preemptive discipline.

The ENo discipline performance is normally very close to non-preemptive scheme performance in small scenarios because the probability of an HP arrival in the first recovery period is very low. In large scenarios, their performance difference increases.

In Figure 10, arrival rates are assumed constant and equal ($\lambda_1 = \lambda_2 = 0.03$) and the system time of both packet types are shown versus the variations of the duration of availability periods (large scenario). The lowest value of $E[Y]$ is selected equal to $E[R] = 15$ because we assumed that it does not make sense to use Opportunistic Spectrum Access (OSA) when availability periods are shorter than the recovery periods. Figure 11 inversely shows the system time versus the variation of the recovery time when the average of availability periods' duration is assumed equal to $E[75]$.

In Figure 12, the packet lengths are both constant (DetExp). Thus, a lower bound for the waiting time of LP packets in the new preemptive scheme has been presented. Moreover, the results are generated for two different values of HP arrival rate. To see the accuracy of completion time approximations, simulation results and approximations for the moments of the completion time of LP packets, D_2, are compared in the upper part of Table I. Note that the completion time of LP packets is independent of their arrival rate.

B. Constant Recovery Time

Constant recovery time indicates, for instance, the scenarios in which the information of channels' occupancy is provided

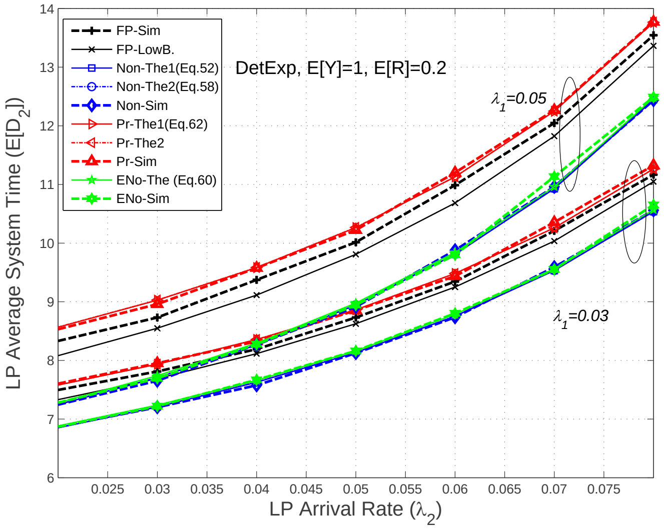

Fig. 12. System time of LP packets vs. LP arrival rate for two values of HP arrival rate (small scenario).

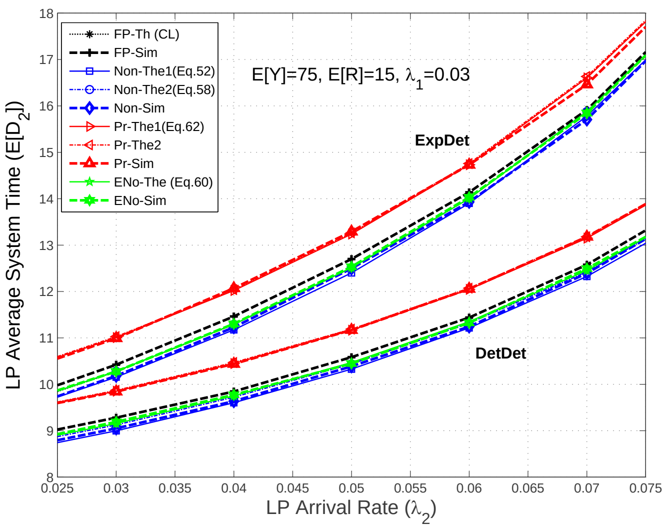

Fig. 13. System time of LP packets vs. LP arrival rate (ExpDet, large scenario).

in advance; therefore, no random sensing is required and the recovery time only represents a constant time for negotiation and radio alignment. In order to compare the results with the previous scenario, we assume the same average. The remaining time after an arrival in R will be distributed by $R_r = R - A|(R > A)$.

When real service time (packet length) is Exponential (ExpDet), waiting time of HP and LP packets in the FP discipline can be found. Figure 13 illustrates the system time of LP packets versus their arrival rate. The system time of HP packets in the same scenario is sketched in Figure 14. In the last scenario where it is assumed that the packet length is also constant (DetDet), and the results are shown in Figures 13 and 14, respectively for LP and HP packets. It should be noted that the upper bound for HP packets in this case is loose, so it is not sketched to enhance the clarity of the figure. As expected, it can be seen in both figures that the performance,

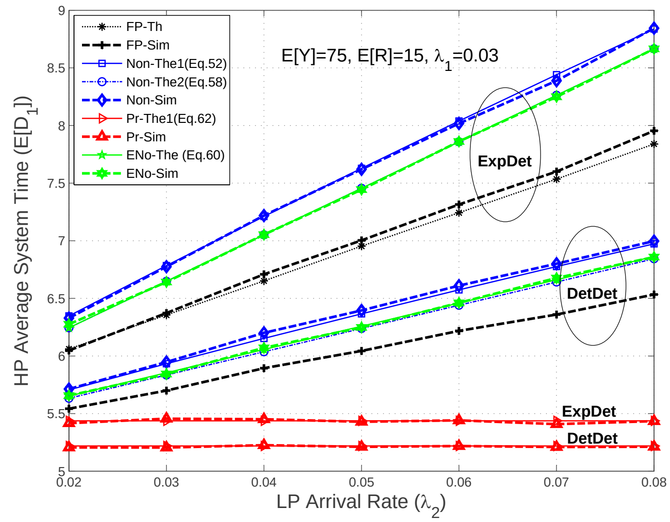

Fig. 14. System time of HP packets vs. LP arrival rate (DetDet and ExpDet, large scenario).

TABLE I
MOMENTS OF THE LP COMPLETION TIME WITH THE POLICY OF PREEMPTION IN CASE OF FAILURE (FP) FOR DETEXP AND DETDET SCENARIOS (S: SIMULATION, A: APPROXIMATION).

Scen.	λ_1	$E[X_2]$-**S**	$E[X_2]$-**A**	$E[X_2^2]$-**S**	$E[X_2^2]$-**A**
Small,DE	0.03	6.62	6.72	47.01	49.51
Small,DE	0.05	7.10	7.32	57.05	62.33
Large,DE	0.03	6.13	6.12	77.53	74.96
Large,DE	0.05	6.26	6.25	88.37	85.62
Small,DD	0.03	6.61	6.72	46.76	49.23
Small,DD	0.05	7.10	7.31	56.69	61.98
Large,DD	0.03	6.14	6.12	58.40	56.45
Large,DD	0.05	6.26	6.25	65.00	63.64

when the distribution of real service time (packet length) is Exponential, is worse compared to the case where the packet length is constant (with the same average).

Again to see the accuracy of approximations for the completion time of LP packets, simulation and analytical results are compared in the lower part of Table I, for two different values of HP arrival rate.

V. CONCLUSION AND FUTURE WORK

Priority queueing is a classical scheme to provide traffic differentiation in communication links. To tackle various priority queueing schemes for opportunistic spectrum access (OSA), implemented by cognitive radios, we discussed in this paper first a general queueing model with interruptions. The derived results were used to mathematically solve three different priority disciplines in the presence of interruptions: non-preemptive, exceptional non-preemptive and preemptive-resume, and to discuss and provide some approximations for a new priority scheme which is preemptive in case of failure. By simulation, derived results and approximations were validated for different scenarios.

As discussed in the introduction, one application of these derived results can be to use them in an optimization and

decision-making problem in cognitive radio networks where the cognitive radio node is able to change the priority queueing scheme on the fly, and the objective is optimizing some performance metrics for different classes of traffic. In such a decision-making model, the decision variable is the priority queueing discipline to be selected among these proposed schemes. This implies that for instance when the number of accumulated LP packets increases, the CR node may start to employ the new scheme of preemption in case of failure, instead of a pure preemptive. Derived results is used to find the immediate cost of making a decision. This idea will be explored in our future work. Moreover, queueing with service repeat after an interruption and with variable service rate will be also discussed in future.

APPENDIX

For two arbitrary random variables X and Y, the distribution of two new random variables, derived from X and Y, are used several times throughout the paper: $Z = X|(X < Y)$ and $Q = X - Y|(X > Y)$. For simplicity of presentation, we discuss the statistics of these two random variables here. We have:

$$f_Z(t) = \frac{Pr(Y > t)f_X(t)}{Pr(Y > X)} = \frac{(1 - F_Y(t))f_X(t)}{Pr(Y > X)}. \quad (87)$$

When Y is exponentially distributed with parameter α, we have:

$$Pr(X < Y) = \int_0^\infty e^{-\alpha t} f_X(t) = \hat{X}(\alpha), \quad (88)$$

$$f_Z(t) = \frac{e^{-\alpha t} f_X(t)}{\hat{X}(\alpha)}. \quad (89)$$

In this case, $E[Z]$ can be given by:

$$E[Z] = \frac{-d/d\alpha \hat{X}(\alpha)}{\hat{X}(\alpha)}. \quad (90)$$

The second moment can be given similarly:

$$E[Z^2] = \frac{d2/d\alpha^2 \hat{X}(\alpha)}{\hat{X}(\alpha)}. \quad (91)$$

For the second random variable, Q, we still assume that Y is Exponentially distributed with parameter α. Then, by doing algebra (details can be found in [6, Lemma2]) or in [8], we have:

$$E[Q] = \frac{E[X]}{1 - \hat{X}(\alpha)} - \frac{1}{\alpha}, \quad (92)$$

and

$$E[Q^2] = \frac{E[X^2] - 2\frac{E[X]}{\alpha}}{1 - \hat{X}(\alpha)} + \frac{2}{\alpha^2}. \quad (93)$$

REFERENCES

[1] F. Jondral, "Cognitive radio: A communications engineering view," *IEEE Wireless Communications*, vol. 14, no. 4, pp. 28–33, Aug. 2007.

[2] J. Mitola, "Cognitive radio architecture evolution," *Proceedings of the IEEE*, vol. 97, no. 4, pp. 626–641, Apr. 2009.

[3] A. Azarfar, J.-F. Frigon, and B. Sansò, "Analysis of cognitive radio networks based on a queueing model with server interruptions," Submitted to IEEE ICC 2012, Sep. 2011.

[4] M. Rashid, J. Hossain, E. Hossain, and V. Bhargava, "Opportunistic spectrum access in cognitive radio networks: A queueing analytic model and admission controller design," in *Proc. of IEEE GLOBECOM '07.*, Nov. 2007, pp. 4647 –4652.

[5] A. Laourine, S. Chen, and L. Tong, "Queuing analysis in multichannel cognitive spectrum access: A large deviation approach," in *Proc. of IEEE INFOCOM 2010*, Mar. 2010, pp. 1 –9.

[6] A. Federgruen and L. Green, "Queueing systems with service interruptions," *Operations Research*, vol. 34, no. 5, pp. 752–768, 1986.

[7] B. Avi-Itzhak, "Preemptive repeat priority queues as a special case of the multipurpose server problemII," *Operations Research*, vol. 11, no. 4, pp. 610–619, 1963. [Online]. Available: http://or.journal.informs.org/content/11/4/610.abstract

[8] H. Takagi, *Queueing analysis: a foundation of performance evaluation, Volume 1*, ser. North-Holland. North-Holland, 1991.

[9] J. D. P. Gaver, "A waiting line with interrupted service, including priorities," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 24, no. 1, pp. 73–90, 1962.

[10] D. Fiems, T. Maertens, and H. Bruneel, "Queueing systems with different types of server interruptions," *European Journal of Operational Research*, vol. 188, no. 3, pp. 838–845, Aug. 2008.

[11] A. Azarfar, J.-F. Frigon, and B. Sansò, "Improving the reliability of wireless networks using cognitive radios," *IEEE Communications Surveys and Tutorials*, vol. 99, pp. 1–17, Mar. 2011.

[12] D. R. Cox, *Renewal Theory*. Methuen; Wiley London, New York, 1962.

[13] S. M. Ross, *Introduction to Probability Models, Ninth Edition*. Orlando, FL, USA: Academic Press, Inc., 2006.

[14] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Prentice-Hall, 1992.

[15] L. Kleinrock, *Queuing Systems*. Wiley, 1975.

[16] H. Khalife, N. Malouch, and S. Fdida, "Multihop cognitive radio networks: to route or not to route," *IEEE Network*, vol. 23, no. 4, pp. 20–25, Jul. 2009.

TABLE II
NOTATIONS

Notation	Description
Y	Length of availability periods (RV)
R	Length of recovery (interruption) periods (RV)
C	C=Y+R
λ	Arrival rate
A	Inter-arrival time
X	Completion time (RV)
Index $_b$	For packets entered a busy system
Index $_e$	For packets entered an empty system
Index $_a$	For packets entered an empty-available system
Index $_{*u}$	For packets entered an empty-unavailable system
Index $_u$	For packets whose service started at the beginning of a Y
B	HP busy periods (RV)
B_b	Busy period started with X_b
B_Z	Busy period started with $Z+X_b$
T	Real service time (RV)
O	$Min(C,T_2)$ (RV)
J_C	Remaining completion time of the packet in service
J_S	Remaining service time of the packet in service
$\hat{Z}(s)$	LST of a continuous random variable Z
$f_Z(t)$	P.D.F of a random variable Z
$F_Z(t)$	C.D.F of a random variable Z
$m(t)$	Average number of renewals until time t
$m_{a\|b\|u}(t)$	m(t) for packets of type a, b or u
$m^2(t)$	Second moment of the number of renewals until time t
ρ^*	$\lambda E[X^*]$
ρ	$\lambda E[X_b]$
T_R	Remaining transmission time in the last period (no interruption)
P_a	Probability of system being available $=\frac{E[Y]}{E[Y]+E[R]}$
P_{ae}	Probability of system being available when empty (arrival point)
W	Waiting time in the queue
D	Total time spent in the system
N^Q	**Average** number of packets waiting in the queue
$P_{aiC\|Y\|R}$	Probability of HP arrival in a cycle (C), Y or R
K	Local parameter to count the number of an event
$i=1,2$	Superindex or Subindex represents the traffic class
α	Exp. Dist. parameter when $F_Y=1-e^{-\alpha t}$
β	Exp. Dist. parameter when $F_R=1-e^{-\beta t}$
$\tilde{F}_Z()$.	Defined in Eq. (16) as the minimum of random variable Z and an Exponential distribution
γ	Exp. Dist. parameter when $F_T=1-e^{-\gamma t}$
$\chi(t)$	Completion time of a packet with length t over a queue with interruption
N_p	Number of priority classes
T_v	Service time of the virtual packets which form the interruptions
R_r	Remaining of R after an arrival in R
C_r	Remaining of C after an arrival in C
$Z_1^{<Z_2}$	$Z_1\|(Z_1<Z_2)$ (for two RVs)
S	Initial setup time (RV) (Section II-H)
P_0	Probability of system being empty